THE GIANT SYSTEM

A COMPUTER SYSTEM FOR THE GENEALOGIST

TALK DELIVERED

AT

WORLD CONFERENCE ON RECORDS

AUGUST 7, 1969

by

LYALL J. GARDNER, B.A.,M.A.
VICE PRESIDENT
SYSTEMS DEVELOPMENT & PROGRAMMING
MANAGEMENT SYSTEMS COMPANY
SALT LAKE CITY, UTAH
801-363-1511

Approved For Release 2006/09/25 : CIA-RDP73-00402R000100140017-1

OBJECTIVE

The objective of the GIANT System is to collect, process and disseminate genealogical information on all deceased persons. Obviously this cannot be done in one step but will have to be achieved in gradual, well planned phases. So far over 100 man years have gone into the development of this system. The first phase of the GIANT System is designed to collect the basic information about individuals and to store that data in one large file. Later phases will be designed to retrieve the information and to correlate the individual records.

The GIANT System is not a hypothetical dream. We committed ourselves to have the computer programs ready for operation by January 1, 1970. The basic programs are now complete and are working.

Let me take just a minute and tell you how a project as large and complex as the GIANT System was organized to achieve the results that we have been able to accomplish.

First, there was no question in the minds of the Management of the Genealogical Society that changes were necessary and that those changes should include the use of the computer. Therefore, there was no problem of selling management. They were behind the project from the very beginning.

Second, there was no one person who knew all of the intricacies of both the genealogical requirements and the complexities of computer technology.

Therefore, a development team consisting of an experienced genealogist and a computer systems engineer were given the responsibility for development of this System. During the time we have been working together we and our staffs have become very aware of the total complexities of the project.

The fact that the system is now in operation is proof that the organization was workable and that those who assisted us were truly talented and dedicated men and women.

HISTORY

The forerunner of the GIANT System was the Records Tabulation System. This was a system in which parish registers of England were converted into machine language by means of flexowriters. These records were then processed by the computer. The end products were

1. an alphabetic printout of the parish register for the use of the researcher and
2. a magnetic tape of the parish in machine language.

The alphabetic listing was of real assistance to the genealogical researcher. There are now over 1200 of these printouts in the Genealogical Society Library. The magnetic tapes were of little value for research since each parish was maintained on a separate tape. All checking for duplicates had to be done manually, and inputing was restricted as an internal project of the Genealogical Society.

The GIANT System retains all the advantages of the Records Tabulation System. In addition every record in which an individual is sufficiently identified is placed in the Genealogical Mass File. This file can be used for duplicate checking automatically by the computer. Also, the inputing has been broadened so that anyone can input information into the GIANT system and have it processed.

THE GIANT SYSTEM

There were four basic criteria for the design of the GIANT System, the first requirement was that the input must be simple. For we are expecting this system to be used by over two million members of the church. The input must be simple enough for a non-experienced genealogist to use with a very minimum amount of instruction. The other criteria for the design were that the system should be accurate, fast and economical.

Because the field of computer technology is progressing so fast, a basic concept of the system design was that it should be modular so that as new devices or techniques are developed they could be incorporated into the system without a complete redesign. This modularity was achieved by dividing the system into eleven (11) sub-systems. Each sub-system is independent and is linked to other sub-systems by the common format of the data records as they are passed from one sub-system to the next. The modular design was also carried down to the program level. So that as modifications are found necessary they can be made quickly without affecting, the logic of the entire program.

The system was written in the COBOL language with some use of Assembler Language and RPG. There are over 60 programs in the whole system. It is operating in a batch mode on an IBM 360/50 but later this month will be transferred to an IBM 360/65.

For those of you who are not acquainted with this machine let me tell you a little about it. Our model 50 machine has over 500,000 positions of main memory. This means that the machine can get to any piece of information in its memory within a millionth of a second. Or it can perform 250,000 additions in a second. Our new model 65 will be four times as fast as our present machine. Besides this main memory we have random access discs attached to the system. With these discs we can have access to any of 500,000,000 characters of information within 60 thousandths of a second. On our new machine this storage will be increased to about 700,000,000 characters of information. Besides this we have over 4000 reels of magnetic tape containing genealogical information.

Because the machine is so fast it has been programmed to process many jobs simutaneously. Thus, it might be

processing several batches of genealogical work, printing payroll checks, updating general ledgers, writing accounts payable checks, and printing labels for the mailing of magazines all at the same time. This is a complicated machine, and it can do some very fantastic things. That is why in the design of the GIANT system it was possible for us to make the input so simple. Everything that was complex we instructed the machine to do.

THE INPUT SYSTEM

I would now like to describe some of the computer aspects of the GIANT system. In order to get information into the computer for processing it must be processed by the input sub-system. There are three basic types of input data or batch types as we call them. These are:

1. Controlled Extraction - This type contains complete recording of parish register or similiar sources of information. Generally these will be christenings, marriages, and burial records.

2. Patron Input - This type of input is the type that anyone of you might insert into the file. It need not be a complete register or a complete family. You can submit as little as one name if you wish.

3. Conversion Batches - Within the Genealogical Society there are many files that must be converted eventually. There are over 36 million names on 3 x 5 index cards and 7 million family group sheets. All of these will eventually have to be put into the Genealogical Mass File.

Our current method of input is by means of punched paper tape. A batch of data to be entered is given to a flexowriter operator. She types the data producing a punched paper tape. The input batch is then given to a second flexowriter operator and she types the same data again and produces a second punched paper tape. The two paper tapes are then compared on the computer and discrepencies printed out for manual evaluation. Besides discrepancies this program also makes checks to insure that names are not put where dates should be, etc. Corrections are made until the input is free from errors and considered "clean". Once the information is "clean" a copy is made and deposited in the Granite Mountain Records Vault.

THE LOCALITY SYSTEM

The "clean" input is next processed through the locality sub-system. The purpose of this sub-system is to code localities such that like localities are combined.

For instance, Constantinople and Istanbul would be filed together. This coding is achieved by means of passing every record against a locality catalog in which localities and their geographic coordinates are found. The input record then contains both the original locality as input, and the geographic coordinates which will be used by the computer to file this record. Besides the coordinates we also keep a country and county code as a part of the total location code.

The coordinates are used only for storing and comparing purposes. They will never be printed out to the genealogical researcher. The true name of the location will be used. For those of you who live in Salt Lake you will not have to remember that Salt Lake is 111 W and 40 N. No locality abbreviations are required. You can submit localities written out completely or you may use standard abbreviations.

A unique computer application was used in connection with the development of the locality catalog. We were unable to locate a gazateer in machine language which had (1) Town, (2) County, (3) Country, and (4) Geographic coordinates. We were able to find, however, one which had everything with the exception of county. Therefore, we developed a routine for assigning the county automatically. On the first country in which this technique was used, we assigned automatically approximately 83 percent of the towns into their correct counties.

If an input record contains a location that is not included in our catalog then that place is added. Thus, the catalog will continue to grow until every inhabited location on earth is included. This catalog is maintained on discs so that locations can be accessed randomly.

THE NAME SYSTEM

After the locality codes are assigned, every given and surname is matched against a name catalog. The purpose of this step is to combine names which are really the same and file them under a standard spelling. Thus, William and Bill will be filed together. All the various spellings of Brown would be treated the same. Languages are kept separate so that "Jon" could be handled differently for French, German, or English. At the present time the name catalog contains over 150,000 given names and 350,000 surnames. The name catalog can also be used to translate names or abbreviations. Thus, the Latin Guilliam can be translated to William or the abbreviation Wm or Eliza. would be translated to William and Elizabeth respectively.

B/C MATCH

The parish register data which enters the system through the Controlled Extraction input system is processed through a Burial/Christening Match program when both burials and christening are present for the same parish. The purpose of this program is to attempt to match the records from these two sources. If a match can be found, then the death or burial date is added to the christening record. The criteria for matching is rather complex and involves the dates, standard spellings of names, locality codes, and relationships as given in the two records. Currently about 10 percent of the christening records are matched with burials by this program.

PUBLIC PRINT

The Controlled Extraction Christening and Marriage type records are processed through a program which sorts, edits, and prints the records in an easy to use form. In original sources it is difficult for the genealogical researcher to find specific information about a person because everything is recorded chronologically and there is generally no index. Our printout is sorted alphabetically by surname and does have a summary of all the names that are found in the parish with their spelling variations.

SEARCH AND UPDATE

The heart of the GIANT System is the Genealogical Mass File. After three years of operation, this file will contain approximately 17 million records. Thereafter, it will grow at the rate of a least three million records per year. This file contains all of the information that is necessary to uniquely identify a person. The file is organized on an individual basis rather than on a family basis. Each new input to the file is checked against the present contents to insure that duplicates are not inserted. If a patron submits a name and the name has already been placed in the file, then he will be notified that the name is already present. The basic items of information that are kept in this file are:

1. The name of the individual.
2. The date of an event such as a birth, death, or marriage.
3. The place in which the event took place.
4. Any relationships to other individuals such parents or spouses.

The records are variable length generally from 100 to 150 characters long and are stored on discs so that they can be searched randomly. The records are filed and searched on uniquely identifying information which is located at the beginning of the record. Many methods of file organization were studied in the design of this mass file. We chose what we believe to be the best method based on the facts that were available. However, there has been built into the system a very elaborate method of gathering statistics on file activity so that by analyzing these figures after some period of operation we will be able to modify the system and make it more efficient.

TIB CHECK LIST AND UPDATE

As the names are matched to the Genealogical Mass File there will be a very high percentage, at first, that do not match. These names will be printed on slips so that they can be manually checked against a card file of 36 million names. If the name is found in the manual file, then additional dates found on the manual card will be added to the computer record. Eventually the complete manual file will be converted and this step will no longer be necessary.

ORDINANCE LISTS AND UPDATING

The last regular step in the GIANT System is the printing and updating of ordinance lists. Here the names of those persons who have not received the ordinances in the temples are printed on lists and sent to the temples to have the ordinances performed vicariously. The dates are inserted into the file and processing is completed.

UTILITIES

Obviously, with a system as large as the GIANT System, there are many utility type programs that are necessary. These programs print, sort, and reorganize files and catalogs for more efficient use, and for copying files for backup and security purposes.

PRODUCTION CONTROL SYSTEM

Earlier I said that each sub-system was independent of every other sub-system. There is one exception, and that is this. Every program affects or is affected by a Production Control System. This is a system which maintains control of the entire system at all times. We estimate that when the system is in full operation after January, 1970, there will be between 8,000 and 10,000 batches of input somewhere in the system at all times. I am sure that you can realize the enormous task of keeping track of where every batch is at all times. We must insure that no batches are lost, that a certain backlog is kept in every phase which requires manual processing so that people don't run out of work, and that certain types of batches are given priority over other types.

Besides controlling the processing of batches through the system, the production control system maintains logs which automatically let us know when certain tapes or discs are no longer needed. The system also keeps statistics on volumes and speeds of processing. These are fed into a Management Information System which gives the Management and operating personnel of the Genealogical Society pertinent information.

FUTURE-PHASE I

This is Phase I as it is today. There are several parts of the system in Phase I that I am sure will be looked at very critically next year. These would include using video terminals to maintain name and locality catalogs, production control files, and to correct input. When optical scanning is perfected and there is time available on our optical scanning machine, we would want to seriously look at changing our input from punched paper tape to straight typing and optical scanning.

FUTURE-PHASE II

The development of PHase II will begin very soon. The prime function of Phase II will be to provide search and retrieval capabilities to the system. These can range from a very simple search where all of the identifying information is known to a very elaborate and complex search where only vague information is known. The degree of search complexity which we design and implement will be dependent on the desires of the Genealogical Society.

Let me elaborate some on these two types of searchs. The first is the simple search in which the identifying information about an individual is known. If a researcher wishes to inquire about a William Taylor, born in Halstead, Essex, England, on the 15th of June 1789, the Genealogical Mass File would be searched for such a record. In order to process this inquiry the names and locality would be checked against the catalogs to determine their codes. If there was a match in the Genealogical Mass File the complete record or records would be printed out for the researcher to examine.

The second type of search is the complex search. In this type of search, the researcher is looking for a person in which he does not know all of the significant identifying information. For example, he may want to know about a child of David Brown and Barbara Green. With present manual techniques, this type of search is completely infeasible. However, with the computer it becomes relatively easy if it is planned for in advance. This we have done. During the processing of the Genealogical Mass File a series of indexes will have been generated. These are lists of identification numbers for items of similar information. Thus, for example there will be a series of indexes for a father's given name.

This index will contain all of the different father's given names that are contained in the Genealogical Mass File. Under each different father's given name will be the identification numbers of the entries that contain these names. This technique will be continued for other items of identifications such as other names, dates, and locations. In order to find the child of David Brown and Barbara Green we program the computer to look for the identification numbers which are common in the indexes of the father's and mother's given and surnames. If there is a common identification number, the complete record is retrieved.

With this technique we could make a search with rather vague paramemeters such as, a child of Paul Thompson and Elizabeth born within 15 miles of Smiths Falls, Ontario, Canada between 1850 and 1880.

We believe that with these two types of searches, a researcher will have at his disposal a tool which will be invaluable as he attempts to do genealogical research.

FUTURE-PHASE III

Phase III of the GIANT System could be the automation of the library. When the library catalog is placed on the computer, we would not only be able to give references to books that are now cataloged but we would also know what will be available later through the microfilming program. If a researcher should ask, "What vital records do you have from Coggeshall, Essex, England?" We would respond with one of four replies:

1. We have the source in the library and give the call number.

2. The source is not in the library but has been microfilmed and should be available by December 17.

3. The source is not in the library but we know that it is found in the public library in New York.

4. We are sorry we have no information about the requested source.

CONCLUSION

In conclusion let me answer a question that I posed at the beginning. "Are you going to be identified in our file simply as a number?" No--all checking is done on information other than on an individual number. There is a number that is used internally by the computer but it is not used outside of machine processing. So you could never find out your number even if you wanted to. Besides, you would have no need to know your number because information about you cannot get into our file until you are dead. And I do not know why you would want to know your number then.

SUMMARY

In summary let me say that it has been a real challenge to be involved in the development of this system. I think it is a good system. It works and it is practical. During the next hour you will be told the details of how it will affect the genealogist. But from our point of view we have met the original cirteria. That is, it is simple, accurate, fast and economical. And most important it works.

I hope that when you leave here this morning you will have learned a little bit about the GIANT System. I hope that you will be aware of what you can do to assist in building this Genealogical Mass File for mankind. But more inportant, I hope that you will have learned what GIANT can do for you and for your ancestors, -- not five years from now, not a year from now, but today.